Math. Program., Ser. A 99: 49–61 (2004)

Digital Object Identifier (DOI) 10.1007/s10107-003-0421-7

Walter F. Mascarenhas

The BFGS method with exact line searches fails for non-convex objective functions

Received: December 25, 2002 / Accepted: March 24, 2003
Published online: May 7, 2003 –

Abstract. This work shows that the BFGS method and other methods in the Broyden class, with exact line searches, may fail for non-convex objective functions.

1. Introduction

Quasi Newton methods revolutionized nonlinear optimization in the 1960's because they avoid costly computations of Hessian matrices and perform well in practice. Several kinds of them have been proposed, but since the 1970's the BFGS method became more and more popular and today it is accepted as the best Quasi Newton method. Along the years, many attempts have been made to find better methods and most of the candidates for "best variable metric method" are members of the so called Broyden class.

The success of the BFGS method and the possibility of success of the other members of the Broyden class have made them the subject of intense scrutiny by theoreticians and practical users of nonlinear optimization. In particular, there are many research works about the convergence of such methods. The references on this subject go as far back as the 1960's [3] and, except for a few minor points, the theory of convergence for convex objective functions is complete. However, a challenging and intriguing question remained open along these forty years: do the BFGS method and the other members of the Broyden class always converge for non-convex objective functions? If not, which members of the family always work and which may fail sometimes? Is there a clever line search strategy which would guarantee convergence? This problem has resisted several attacks. Some, like [5], almost solved it by proposing small changes to the methods. In [7] M.J.D. Powell presented an example with two variables in which the method fails for a particular line search and in [8] he proved convergence when there are only two variables and the line searches find the first local minimizer. Finally, in [1], Yu-Hong Dai showed that the BFGS method may fail for non-convex functions with line searches that satisfy the Wolfe conditions.

This work answers this question when at each step we pick a global minimizer along the search line. We present a geometric mechanism that makes the BFGS method fail in this circumstance. We emphasize that this specific line search strategy is natural and,

W.F. Mascarenhas: Dept. of Computing and Statistics, UFMS, Campo Grande, Brazil,
e-mail: walter@geocad.com.br

according to [2], our example leads to the failure of all members of the Broyden family for which a step length of $\sqrt{2}$ does not create degenerate matrices B_k. We have found similar examples for other line searches, like the one that picks the first minimizer along the search line, but they are rather technical and the example presented here strikes the best balance to deliver our message: the BFGS method may fail for non-convex objective functions when the line searches find a global minimizer along the search line.

We want to show that convexity is important and use strong conditions regarding other factors that affect convergence. In this spirit, the iterates not only are found by exact line searches, but also satisfy an Armijo condition. More precisely, we present an objective function f and iterates x_k that satisfy the Armijo condition

$$f(x_{k+1}) - f(x_k) = \frac{(\sqrt{2}-1)^2}{5}(x_{k+1}-x_k)^T \nabla f(x_k) \approx 0.034(x_{k+1}-x_k)^T \nabla f(x_k) \quad (1)$$

and for which x_{k+1} is the global minimizer of f in the straight line $x_k \to x_{k+1}$, i.e.

$$t \neq 0 \Rightarrow f(x_{k+1}) < f(x_{k+1} + t(x_k - x_{k+1})). \quad (2)$$

The example is remarkably simple and is sketched in Figure 1:

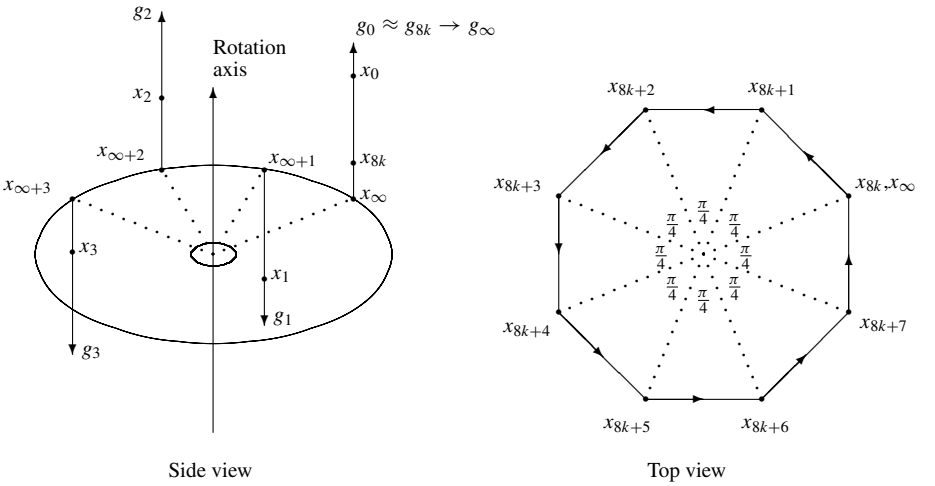

Fig. 1. The example from two perspectives

The iterates x_k are three dimensional vectors. At each iteration the horizontal component rotates counterclockwise by $\pi/4$ around the vertical axis, staying forever at the vertices of a regular octagon. The vertical component converges linearly to 0 at a rate of $1/\sqrt{2}$, flipping sign at each step. Both B_k and its inverse grow without bound. Finally, and more importantly, the gradients oscillate around $\pm(0, 0, 1)^T$. As a result, the iterates never get close to a local minimizer and the method fails.

We present now formal definitions of the iterates in terms of vectors and matrices, state several results about them and leave the proofs of these results for latter sections. The iterates are

$$x_k = (XQ)^k (x_\infty + e_z), \tag{3}$$

where x_∞ is the vertex of the regular octagon in the figure above, e_z is the vertical unit vector and the scaling matrix X is responsible for the decay of the z coordinate of x_k:

$$x_\infty = \frac{1}{2}\begin{pmatrix} 3+2\sqrt{2} \\ 1+\sqrt{2} \\ 0 \end{pmatrix}, \quad e_z = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad X = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2^{-1/2} \end{pmatrix}. \tag{4}$$

The orthogonal matrix

$$Q = \begin{pmatrix} 2^{-1/2} & -2^{-1/2} & 0 \\ 2^{-1/2} & 2^{-1/2} & 0 \\ 0 & 0 & -1 \end{pmatrix} \tag{5}$$

rotates x_∞ by $\pi/4$ in the counterclockwise direction and flips the sign of e_z.

The gradients are

$$g_k = (GQ)^k g_0, \tag{6}$$

for

$$G = \begin{pmatrix} 2^{-1/2} & 0 & 0 \\ 0 & 2^{-1/2} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad g_0 = \begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix}. \tag{7}$$

The Hessian approximations are

$$B_k = -\frac{\sqrt{2}}{s_k^T g_k} g_k g_k^T - \frac{\sqrt{2}}{s_{k+1}^T g_{k+1}} g_{k+1} g_{k+1}^T - \frac{\sqrt{2}}{s_{k+2}^T g_{k+2}} g_{k+2} g_{k+2}^T, \tag{8}$$

for $s_k = x_{k+1} - x_k$. In particular, B_0 is the positive definite matrix

$$B_0 = \frac{\sqrt{2}}{5}\begin{pmatrix} 11 & -7 & 12 \\ -7 & 9 & 6 \\ 12 & 6 & 4 \end{pmatrix} - \frac{1}{5}\begin{pmatrix} 3 & -11 & 16 \\ -11 & 7 & 8 \\ 16 & 8 & 2 \end{pmatrix}.$$

We also have

$$B_k = M^k \{Q^k B_0 (Q^T)^k\} M^k \tag{9}$$

where the scaling matrix

$$M = 2^{1/4} G = \begin{pmatrix} 2^{-1/4} & 0 & 0 \\ 0 & 2^{-1/4} & 0 \\ 0 & 0 & 2^{1/4} \end{pmatrix} \tag{10}$$

is responsible for the unlimited growth of B_k and its inverse. The reader can verify (9) noticing that G, M and X are diagonal matrices that commute with Q and using equations (6)–(8) and (11).

The step length α_k, for which $B_k s_k = -\alpha_k g_k$, is constant and equals $\sqrt{2}$ and the properties of x_k, g_k and B_k are summarized in the following theorem:

Theorem 1. *The step length $\alpha_k = \sqrt{2}$, x_k, g_k and B_k defined by (3) – (8) satisfy*

$$s_k^T g_k = -\frac{5}{2-\sqrt{2}}\, 2^{-k/2}, \tag{11}$$

the iteration equations for the BFGS method

$$B_k s_k = -\alpha_k g_k, \tag{12}$$

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{s_k^T y_k}, \tag{13}$$

for $y_k = g_{k+1} - g_k$, and the condition

$$s_k^T g_{k+1} = 0, \tag{14}$$

which in necessary for exact line searches.

Now we need a function f that generates these iterates. Our candidates are defined locally around the vertices $Q^k x_\infty$ of the octagon as the cubic function

$$q_k(w + Q^k x_\infty) = ((-1)^k w_3)\left(1 + w^T Q^k h + 1.1(w^T Q^k d)^2\right), \tag{15}$$

where $h = (3, -1, 0)^T$ is the horizontal component of g_0 and $d = (0, 1, 0)^T$. Since Q^k has period 8 in k there are only 8 distinct functions q_k and we have the theorem:

Theorem 2. *The function q_k in (15) satisfies*

$$q_k(x_k) = 2^{-k/2} \tag{16}$$

and interpolates the gradient g_k at x_k, i.e., $\nabla q_k(x_k) = g_k$.

As the reader can verify while reading the proof of Theorem 2, as $k \to \infty$ the eigenvalues of the Hessian of q_k at x_k approach 0 and $\pm\sqrt{10}$ with eigenvectors

$$Q^k \begin{pmatrix} 1 \\ 3 \\ 0 \end{pmatrix}, \quad Q^k \begin{pmatrix} 3 \\ -1 \\ -\sqrt{10} \end{pmatrix}, \quad Q^k \begin{pmatrix} 3 \\ -1 \\ \sqrt{10} \end{pmatrix}.$$

Thus the cubic functions q_k are not convex. However, as the next theorem shows, any function that is equal to them near the octagon's vertices has x_{k+1} as a local minimizer along the line $x_k \to x_{k+1}$:

Theorem 3. *If x is a point on the line $x_k \to x_{k+1}$ that satisfies $0 < \|x - x_{k+1}\|_2 < 0.08$, then $q_{k+1}(x) > q_{k+1}(x_{k+1})$.*

The Armijo condition (1) follows from (11) and (16). Therefore, theorems 1, 2 and 3 already show that the BFGS method may fail for line searches satisfying this condition and for which x_{k+1} is a local minimizer of f along the line $x_k \rightarrow x_{k+1}$: take the x_0 and B_0 above and apply the method to any function that matches the cubic functions q_k in the neighborhood of the octagon's vertices, using the step length $\alpha = \sqrt{2}$ for all iterations. However, we go a step further and provide a C^∞ function f that coincides with the q_k's near the octagon's vertices and for which x_{k+1} is a global minimizer along the line $x_k \rightarrow x_{k+1}$.

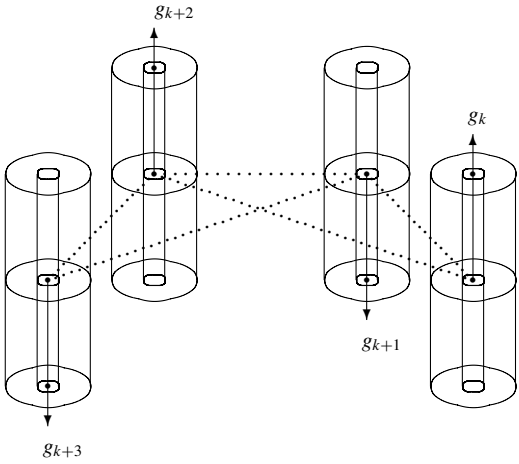

Fig. 2. Inner and outer cylinders around the octagon's vertices

The idea is to take a function f that is defined on each of the pieces of $\Re^3$ described in Figure 2. In the interior of the thin cylinders of radius 0.04 around each vertex of the octagon f equals the cubic function (15). Outside the cylinders of radius 0.16 around the thinner cylinders f is constant and equals 2. Between the interior of the thinner cylinders and the outside of the wider cylinders f goes through a smooth transition from the cubics to the constant function.

More formally, we write x as (c, z), for $c \in \Re^2$ and $z \in \Re$, and take

$$f(c, z) = \sum_{j=0}^{j=7} \psi(c - c_j) q_j(c, z) + 2 \prod_{j=0}^{j=7} (1 - \psi(2(c - c_j))), \tag{17}$$

where c_j is the jth octagon's vertex and $\psi : \Re^2 \rightarrow \Re$ is a C^∞ function such that

$$0 = \inf_{c \in \Re^2} \psi(c) < \sup_{c \in \Re^2} \psi(c) = 1, \tag{18}$$

$$\| c \| \leq 0.08 \Rightarrow \psi(c) = 1, \tag{19}$$

$$\| c \| \geq 0.16 \Rightarrow \psi(c) = 0. \tag{20}$$

This kind of cutoff function is described in page 25 of [4]. A reader not comfortable with them could obtain a C^2 function ψ with the same properties taking $\psi = \sigma(c^T c)$ where $\sigma : \Re \to \Re$ is an appropriate cubic spline.

Along the line from x_k to x_{k+1} f has the graph below. According to this graph, f has two valleys. The higher valley lies around x_k and the lower one has its bottom at x_{k+1}, which is a global minimizer along this line. Notice that f decreases at x_k but we made the cylinder's radius so small that x enters in the transition region and f starts to grow before getting too close to $f(x_{k+1})$.

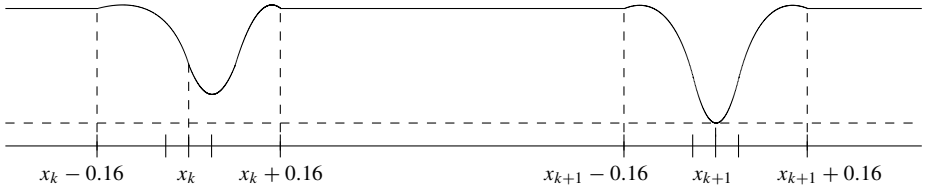

Fig. 3. The graph of f along the line $x_k \to x_{k+1}$

This description of f is summarized in the following theorem, which, with theorems 1 and 2, shows that f fulfils the claims made in (1) and (2).

Theorem 4. *The function f defined by (17) satisfies (2) and*

$$\|c - c_k\| \leq 0.04 \Rightarrow f(c, z) = q_k(c, z). \tag{21}$$

Theorems 1 and 2 are mostly algebraic. Proving them is just a question of going through the algebra, computing a few 3×3 matrix vector products and derivatives. We do that in Section 2. Theorems and 3 and 4 involve estimates and we prove them in Section 3. The proofs are of little interest for the practical minded reader, who is probably asking by now "and what happens in finite precision arithmetic?" A complete answer to this question requires a discussion of the various details involved in "professional" implementations and is out of the scope of this article (and beyond the author's expertise.) We would like, however, to comment that in our naive numerical experiments the behavior suggested by the exact arithmetic analysis is accurate for 20 iterations, or two and a half laps around the octagon. The effects of rounding began at the machine precision level and are amplified by a factor of roughly 10 at each iteration. By the middle of the third lap the rounding errors have taken their toll and the iterates leave the thin cylinders from Figure 2. From then on our analysis is just not valid.

Finally, we would like to thank Prof. J.M. Martinez for bringing the problem solved in this paper to our attention and for the many suggestions that, we hope, have added a nonlinear optimization content to our work and turned our findings into something more valuable than a clever solution to a math puzzle. We also would like to thank Prof. M.J.D. Powell for reading a previous version of this work and making comments that lead to this improved version. Of course, we are solely responsible for our writing style and the other problems they may not have noticed.

2. Algebra

In this section we prove Theorems 1 and 2, in this order.

Proof of Theorem 1. We start using (3) to rewrite s_k as

$$s_k = x_{k+1} - x_k = (XQ)^k(XQ - I)(x_\infty + e_z) = (XQ)^k s_0, \tag{22}$$

for

$$s_0 = (XQ - I)(x_\infty + e_z) = \frac{1+\sqrt{2}}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \\ -1 \end{pmatrix}. \tag{23}$$

The matrices X, Q and G, defined in (4), (5) and (7), commute, $Q^T Q = I$ and $XG = 2^{-1/2} I$, where I is the 3×3 identity matrix. Therefore,

$$s_k^T g_{k+i} = s_0^T ((XQ)^T)^k (QG)^k g_i = s_0^T (XG)^k g_i = 2^{-k/2} s_0^T g_i.$$

In particular, for $i = 0$, 1, 2 and 3,

$$s_k^T g_k = 2^{-k/2} s_0^T g_0 = -\frac{5}{\sqrt{2}}(1+\sqrt{2})2^{-k/2}, \tag{24}$$

$$s_k^T g_{k+1} = 2^{-k/2} s_0^T g_1 = 0, \tag{25}$$

$$s_k^T g_{k+2} = 2^{-k/2} s_0^T g_2 = 0, \tag{26}$$

$$s_k^T g_{k+3} = 2^{-k/2} s_0^T g_3 = \frac{5}{2\sqrt{2}}(1+\sqrt{2})2^{-k/2}, \tag{27}$$

as it can be seen from the definition of Q in (5), G and g_0 in (7), s_0 in (23) and

$$GQ = \frac{1}{2} \begin{pmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}, \quad g_0 = \begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix} \tag{28}$$

$$g_1 = GQg_0 = \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}, \quad g_2 = GQg_1 = \frac{1}{2}\begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}, \quad g_3 = GQg_2 = \frac{1}{2}\begin{pmatrix} -1 \\ 2 \\ -2 \end{pmatrix}. \tag{29}$$

The equations (25) and (26) and the definition of B_k in (8) imply

$$B_k s_k = -\sqrt{2} g_k \tag{30}$$

and the iterates satisfy equation (12) with $\alpha_k = \sqrt{2}$. Since (24) is the same as (11) and (25) is the same as (14), we only need to verify (13) in order to prove Theorem 1.

Equations (28) and (29) lead to $g_1 - g_0 = 2g_3$. Since $y_k = g_{k+1} - g_k$, (6) implies

$$y_k = (QG)^k (g_1 - g_0) = 2(QG)^k g_3 = 2g_{k+3}. \tag{31}$$

Using (27) and (31) and replacing k by $k+3$ in (24) we get

$$s_k^T y_k = 2 s_k^T g_{k+3} = \frac{5}{\sqrt{2}}(1+\sqrt{2})2^{-k/2} = 2^{3/2}\frac{5}{\sqrt{2}}(1+\sqrt{2})2^{-(k+3)/2} = -2\sqrt{2} s_{k+3}^T g_{k+3}.$$

Combining the last equation with (31) we obtain

$$\frac{y_k y_k^T}{s_k^T y_k} = \frac{4 g_{k+3} g_{k+3}^T}{-2\sqrt{2} s_{k+3}^T g_{k+3}} = -\frac{\sqrt{2}}{s_{k+3}^T g_{k+3}} g_{k+3} g_{k+3}^T.$$

Equation (30) leads to

$$\frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} = \frac{2 g_k g_k^T}{-\sqrt{2} s_k^T g_k} = -\frac{\sqrt{2}}{s_k^T g_k} g_k g_k^T.$$

From the last two equations and the definition (8) of B_k we get that the right hand side of (13) is

$$\begin{aligned} &-\frac{\sqrt{2}}{s_k^T g_k} g_k g_k^T - \frac{\sqrt{2}}{s_{k+1}^T g_{k+1}} g_{k+1} g_{k+1}^T - \frac{\sqrt{2}}{s_{k+2}^T g_{k+2}} g_{k+2} g_{k+2}^T \\ &+\frac{\sqrt{2}}{s_k^T g_k} g_k g_k^T - \frac{\sqrt{2}}{s_{k+3}^T g_{k+3}} g_{k+3} g_{k+3}^T. \end{aligned}$$

The first and fourth terms of this sum cancel out and we are left with

$$-\frac{\sqrt{2}}{s_{k+1}^T g_{k+1}} g_{k+1} g_{k+1}^T - \frac{\sqrt{2}}{s_{k+2}^T g_{k+2}} g_{k+2} g_{k+2}^T - \frac{\sqrt{2}}{s_{k+3}^T g_{k+3}} g_{k+3} g_{k+3}^T,$$

which, according to (8), is the expression for B_{k+1}. Therefore, the iterates satisfy equation (13) and the proof of Theorem 1 is complete.

Proof of Theorem 2. The change of variables $w = Q^k u$ simplifies the expression (15) and, since $Q^k e_z = (-1)^k e_z$, leads to

$$q_k(w + Q^k x_\infty) = r(u) = u_3 \left(1 + 3u_1 - u_2 + 1.1 u_2^2\right). \tag{32}$$

The gradient of r is

$$\nabla r(u) = \begin{pmatrix} 3u_3 \\ u_3(2.2u_2 - 1) \\ 1 + 3u_1 - u_2 + 1.1u_2^2 \end{pmatrix}$$

and the chain rule shows that $\nabla q_k(w + Q^k x_\infty) = Q^k \nabla r((Q^k)^T w) = Q^k \nabla r(u)$.

According to (3) and (4), x_k corresponds to $w_k = (0, 0, (-1)^k 2^{-k/2})^T$ and $u_k = (Q^k)^T w_k = (0, 0, 2^{-k/2})^T$, so (32) implies (16). Using the definitions of g_k, g_0 and G, given in (6) and (7), we get

$$\nabla q_k(x_k) = Q^k \nabla r(u_k) = Q^k \left(2^{-k/2} 3, -2^{-k/2}, 1\right)^T = (QG)^k (3, -1, 1)^T = g_k.$$

Thus, q_k interpolates the gradients g_k at x_k and the proof of Theorem 2 is complete.

3. Estimates

In this section we prove Theorems 3 and 4. We think of x as a pair (c, z), where $c \in \Re^2$ and $z \in \Re$. The kth vertex of the octagon is called c_k and is the projection of x_k in the horizontal plane. The expression for x_∞ in (4) leads to

$$c_0 = \frac{1}{2}\begin{pmatrix} 3+2\sqrt{2} \\ 1+\sqrt{2} \end{pmatrix} \tag{33}$$

and (3) and (5) imply

$$c_k = R^k c_0, \tag{34}$$

where

$$R = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \tag{35}$$

is the rotation by $\pi/4$ in the counterclockwise direction. Therefore, the side of the octagon in Figure 1 has length $\|c_1 - c_0\|_2 = 1 + \sqrt{2}$ and

$$k < j \leq k+7 \Rightarrow \|c_j - c_k\|_2 \geq 1 + \sqrt{2}. \tag{36}$$

In terms of R, the cubic function q_k from (15) can be written as

$$q_k(c, z) = z(-1)^k \left(1 + (c - c_k)^T R^k g_c + 1.1 \left((c - c_k)^T R^k d_c\right)^2\right) \tag{37}$$

where

$$g_c = \begin{pmatrix} 3 \\ -1 \end{pmatrix} \quad d_c = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{38}$$

are the projections of h and d in $\Re^2$, because

$$w^T Q^k h = ((c - c_k)^T, 0)\, Q^k \begin{pmatrix} g_c \\ 0 \end{pmatrix} = (c - c_k)^T R^k g_c$$

and

$$w^T Q^k d = ((c - c_k)^T, 0)\, Q^k \begin{pmatrix} d \\ 0 \end{pmatrix} = (c - c_k)^T R^k d_c.$$

Now we prove a slightly stronger version of Theorem 3:

Theorem 5. *If x is a point on the line $x_k \to x_{k+1}$ that satisfies $0 < \|c - c_{k+1}\|_2 < 0.08$, then $q_{k+1}(x) > q_{k+1}(x_{k+1})$.*

This theorem implies Theorem 3 because

$$\|x - x_{k+1}\|_2 < 0.08 \Rightarrow \|c - c_{k+1}\|_2 < 0.08$$

in general and $c = c_{k+1} \Rightarrow x = x_{k+1}$ if x is on the straight line $x_k \to x_{k+1}$.

After that we prove Theorem 4 and finish this section.

Proof of Theorem 5. Let $x = x_{k+1} + \epsilon(x_k - x_{k+1})$ be a point on the line $x_k \to x_{k+1}$. Writing x as (c, z) as in section 2, we get from (34) that

$$c = c_{k+1} + \epsilon(c_k - c_{k+1}) = c_{k+1} + \epsilon R^k (I - R)c_0 = c_{k+1} + \epsilon R^k v$$

or

$$c - c_{k+1} = \epsilon R^k v, \tag{39}$$

for (c_0 is defined in (33) and R in (35))

$$v = (I - R)c_0 = \frac{1 + \sqrt{2}}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}. \tag{40}$$

Using (3) we deduce that $z_k = (-1)^k 2^{-k/2}$ and

$$z = z_{k+1} + \epsilon(z_k - z_{k+1}) = (-1)^{k+1} 2^{-(k+1)/2}(1 - \epsilon(1 + \sqrt{2})).$$

This equation for z, the definition of R in (35), $(R^k)^T R^k = I$, (37)(with k replaced by $k + 1$) and (39) imply

$$q_{k+1}(c, z) = 2^{-(k+1)/2}(1 - \epsilon(1 + \sqrt{2}))\left(1 + \epsilon v^T R g_c + 1.1\epsilon^2 (v^T R d_c)^2\right) \tag{41}$$

Equations (35) and (40) imply $v^T R = (1 + \sqrt{2})(0, -1)^T$ and (38) and (41) lead to

$$q_{k+1}(c, z) = 2^{-(k+1)/2}(1 - \epsilon(1 + \sqrt{2}))\left(1 + \epsilon(1 + \sqrt{2}) + 1.1\epsilon^2(1 + \sqrt{2})^2\right)$$

$$= 2^{-(k+1)/2}\left(1 + 0.1\epsilon^2(1 + \sqrt{2})^2 - 1.1\epsilon^3(1 + \sqrt{2})^3\right).$$

In particular, $\epsilon = 0$ gives $q_{k+1}(x_{k+1}) = 2^{-(k+1)/2}$, in agreement with equation (16). Therefore,

$$q_{k+1}(c, z) = q_{k+1}(x_{k+1})\left(1 + 0.1\epsilon^2(1 + \sqrt{2})^2(1 - 11\epsilon(1 + \sqrt{2}))\right). \tag{42}$$

Since the matrix (35) is orthogonal, and since $\|c - c_{k+1}\|_2 < 0.08$ by hypothesis, equations (39) and (40) lead to

$$(1 + \sqrt{2})|\epsilon| = |\epsilon|\|v\|_2 = |\epsilon|\|R^k v\|_2 = \|c - c_{k+1}\|_2 < 0.08. \tag{43}$$

Therefore, $1 - 11\epsilon(1 + \sqrt{2})$ is bounded below by $1 - 0.88 = 0.12$, so (42) implies the inequality

$$q_{k+1}(c, z) \geq q_{k+1}(x_{k+1})\left(1 + 0.012\epsilon^2(1 + \sqrt{2})^2\right),$$

completing the proof of Theorem 5.

Proof of Theorem 4. Since q_k has period 8 in k, we can extend the definition of c_k for $k \geq 8$ as $c_k = c_{(k \bmod 8)}$ and rewrite f as

$$f(c,z) = \sum_{j=k}^{j=k+7} \psi(c-c_j) q_j(c,z) + 2 \prod_{j=k}^{j=k+7} (1-\psi(2(c-c_j))). \tag{44}$$

If $\| c-c_k \|_2 \leq 0.04$ then (36) and the triangle inequality show that $\| c-c_j \|_2 > 0.16$ for all vertices c_j with $k < j \leq k+7$. Thus, (20) implies $\psi(c-c_j) = 0$ for such j and the sum in (44) equals its first term, $\psi(c-c_k) q_k(c,z)$. Since (19) and $\| c-c_k \|_2 \leq 0.04$ imply $\psi(c-c_k) = \psi(2(c-c_k)) = 1$, the product in (44) vanishes and $f(c,z)$ reduces to $q_k(c,z)$. Thus we have proved (21) and (16) implies

$$f(x_{k+1}) = q_{k+1}(x_{k+1}) = 2^{-(k+1)/2} < 2. \tag{45}$$

In order to verify the property (2) and complete the proof of Theorem 4, we must analyze f along the straight line $x_k \to x_{k+1}$. Elementary geometry and (36) show that the smallest distance from the straight line $c_k \to c_{k+1}$ to the remaining vertices of the octagon is $(1+\sqrt{2})\sin(\pi/4) \approx 1.7 > 0.16$. Therefore, (20) and (44) imply that the function f on the straight line $x_k \to x_{k+1}$ reduces to

$$f(c,z) = \sum_{j=k}^{j=k+1} \psi(c-c_j) q_j(c,z) + 2 \prod_{j=k}^{j=k+1} (1-\psi(2(c-c_j))). \tag{46}$$

The vector $x = (c,z)$ is covered by one of the three cases:

- Case 1: $\| c-c_k \| \geq 0.16$ and $\| c-c_{k+1} \| \geq 0.16$.
- Case 2: $\| c-c_{k+1} \| < 0.16$.
- Case 3: $\| c-c_k \| < 0.16$,

We now analyze the three cases above individually.

Case 1. In this case (20) implies that both terms in the sum in (46) vanish and the factors in the product are 1. Therefore, $f(x) = 2 > f(x_{k+1})$, by (45), and Theorem 4 holds in case 1 •

Case 2. In this case the triangle inequality and $\| c_k - c_{k+1} \|_2 \geq 1+\sqrt{2}$ (see (36)) lead to $\| c-c_k \|_2 \geq 0.16$. Therefore, (20) implies $\psi(c-c_k) = \psi(2(c-c_k)) = 0$ and (46) becomes

$$f(c,z) = \psi(c-c_{k+1}) q_{k+1}(c,z) + 2(1-\psi(2(c-c_{k+1}))). \tag{47}$$

We now proceed as in the proof of Theorem 5, because equation (42) is valid. Expression (43) includes the equations

$$\|c-c_{k+1}\|_2 = |\epsilon| \|R^k v\|_2 = |\epsilon| \|v\|_2 = (1+\sqrt{2})|\epsilon|,$$

so the hypothesis of Case 2 provides

$$(1+\sqrt{2})|\epsilon| = \|c-c_{k+1}\|_2 < 0.16. \tag{48}$$

Thus this hypothesis and (42) imply

$$q_{k+1}(c,z) \geq q_{k+1}(x_{k+1})(1+0.1 \times (0.16)^2 \times (-0.76)) > 0. \tag{49}$$

If $\|c - c_{k+1}\|_2 < 0.08$ then (2) follows from Theorem 5. On the other hand, $\|c - c_{k+1}\|_2 \geq 0.08$ and (20) imply $\psi(2(c - c_{k+1})) = 0$. Therefore, (18), (49), (45) and (47) imply $f(x) \geq 2 > f(x_{k+1})$ and we are done with case 2 •

Case 3. Analogously to (47) in Case 2, we have

$$f(x) = f(c, z) = \psi(c - c_k)q_k(c, z) + 2(1 - \psi(2(c - c_k))), \tag{50}$$

However, now we write x as $x_k + \epsilon(x_{k+1} - x_k)$ and (34) leads to

$$c = c_k + \epsilon(c_{k+1} - c_k) = c_k + \epsilon R^k(Rc_0 - c_0) = c_k - \epsilon R^k v,$$

for R in (35) and v in (40), or

$$c - c_k = -\epsilon R^k v, \tag{51}$$

and in the present case we have a relation analogous to (48):

$$(1 + \sqrt{2})|\epsilon| = \|c - c_k\|_2 < 0.16. \tag{52}$$

The definition of x_k in (3) shows that $z_k = (-1)^k 2^{-k/2}$ and

$$z = z_k + \epsilon(z_{k+1} - z_k) = (-1)^k 2^{-k/2}(1 - \epsilon\frac{1 + \sqrt{2}}{\sqrt{2}}). \tag{53}$$

The definitions of v in (40) and g_c and d_c in (38) show that $v^T g_c = 2\sqrt{2}(1 + \sqrt{2})$ and $v^T d_c = -(1 + \sqrt{2})/\sqrt{2}$. Since $R^T R = I$, equations (37), (51) and (53) imply

$$q_k(c, z) = 2^{-k/2}(1 - \epsilon\frac{1 + \sqrt{2}}{\sqrt{2}})(1 - 2\sqrt{2}\epsilon(1 + \sqrt{2}) + 0.55\epsilon^2(1 + \sqrt{2})^2)$$

$$= 2^{-k/2}\left(1 - 5\epsilon\frac{1 + \sqrt{2}}{\sqrt{2}} + 2.55\epsilon^2(1 + \sqrt{2})^2 - \frac{0.55}{\sqrt{2}}\epsilon^3(1 + \sqrt{2})^3\right).$$

The bound (52) implies

$$2.55\epsilon^2(1 + \sqrt{2})^2 - \frac{0.55}{\sqrt{2}}\epsilon^3(1 + \sqrt{2})^3 > 0.$$

Therefore,

$$q_k(c, z) > 2^{-k/2}\left(1 - 5\epsilon\frac{1 + \sqrt{2}}{\sqrt{2}}\right) \tag{54}$$

and (52) leads to

$$q_k(c, z) > 2^{-k/2}(1 - \frac{0.8}{\sqrt{2}}) > 0. \tag{55}$$

Again, we consider two sub cases

- Sub Case 3.1: $\|c - c_k\| \geq 0.08$
- Sub Case 3.2: $\|c - c_k\| < 0.08$

separately.

Sub case 3.1. Since we are assuming $\|c-c_k\| \geq 0.08$, (20) implies $\psi(2(c-c_k)) = 0$. Therefore, (18), (55), (45) and (50) lead to $f(x) \geq 2 > f(x_{k+1})$ and we are done with sub case 3.1.

Sub case 3.2. Here $\|c - c_k\|_2 < 0.08$ and (19) implies $\psi(c - c_k) = 1$ while (52) can be strengthened to $\epsilon < 0.08/(1 + \sqrt{2})$. It follows that

$$1 - 5\epsilon\frac{1+\sqrt{2}}{\sqrt{2}} > 1 - \frac{0.4}{\sqrt{2}} > \frac{1}{\sqrt{2}}.$$

Using the value of $f(x_{k+1})$ from (45), equation (47), $\psi(c - c_k) = 1$ and the bound (54) we get

$$f(x) \geq q_k(x) > 2^{-k/2}\frac{1}{\sqrt{2}} = f(x_{k+1})$$

and the sub case 3.2 is finished, as well as case 3 •

This completes the proof of Theorem 4 and the justification of the example.

References

[1] Yu-Hong, D.: Convergence properties of the BFGS Algorithm. SIAM J. Optim **13**(3), 693–701 (2002)

[2] Dixon, L.C.W.: Quasi-Newton Algorithms Generate Identical Points. Math. Program. **2**, 383–387 (1972)

[3] Fletcher, R., Powell, M.J.D.: A rapidly convergent descent method for minimization. Comput. J. **6**, 163–168 (1963)

[4] Hörmander, L.: The Analysis of Linear Partial Differential Operators I, Springer Verlag, 1983

[5] Li, D., Fukushima, M.: On the Global Convergence of BFGS Method for Nonconvex Unconstrained Optimization. SIAM J. Optim. **11**(4), 1054–1064 (2001)

[6] Nocedal, J.: Theory of algorithms for unconstrained optimization. Acta Numerica 199–242 (1992)

[7] Powell, M.J.D.: Nonconvex minimization calculations and the conjugate gradient method. In: D.F. Griffiths, ed., Numerical Analysis, Lecture Notes in Mathematics 1066, Springer Verlag, Berlin, 1984 pp. 122–141

[8] Powell, M.J.D.: On the convergence of the DFP Algorithm for unconstrained optimization when there are only two variables. Math. Program. Ser. B. **87**, 281–301 (2000)